

5 We claim:

- 5b
A' 7
1. An enhanced interactive voice response system, comprising:
a call router to route an internet protocol telephony call; and
an interactive-voice-response server to receive the internet protocol
10 telephony call from the call router, wherein the interactive voice response server
includes a terminal object.
 2. The system of claim 1, further comprising a gateway coupled to the call
router.
 - 15 3. The system of claim 2, further comprising a public switched telephone
network coupled to the gateway.
 4. The system of claim 2, wherein the gateway translates telephony calls
20 based on communication protocols of a public switched telephone network to
telephony calls based on internet protocols.
 5. The system of claim 1, further comprising a client computer, wherein the
client computer includes a terminal object so as to receive the internet telephony
25 call routed from the router.

5

6. An enhanced interactive voice response system, comprising:

a data store;

a call router to route an internet protocol telephony call; and

an interactive voice response server to receive the internet protocol

10 telephony call from the call router, wherein the interactive voice response server includes a terminal object.

7. The system of claim 6, wherein the call router stores call information in the data store.

15

8. The system of claim 6, wherein the interactive voice response server stores call information in the data store.

9. The system of claim 6, further comprising a client computer, wherein the
20 client computer includes a terminal object so as to receive the internet telephony call routed from the router.

10. The system of claim 9, wherein the client computer is adapted to retrieve call information from the data store.

25

5 ~~11.~~ An enhanced unified message system, comprising:
 an email store; and
 a voice mail system to receive an internet protocol telephony call, wherein
the voice mail system includes a terminal object.

10 12. The system of claim 11, further comprising a gateway to transmit an
internet protocol telephony call.

13. The system of claim 12, further comprising a client computer to receive the
internet protocol telephony call from the gateway.

15 14. The system of claim 13, wherein the voice mail system saves the internet
protocol telephony call in the email store.

20 15. The system of claim 14, wherein the client computer is adapted to access a
saved internet protocol telephony call through the email store.

~~16.~~ A system to enhance speech-enabled Web applications, comprising:
 a Web page that includes voice tags; and
 a voice browser that includes a terminal object to interpret the voice tags.

25

5 17. The system of claim 16, further comprising a Web server that stores the Web page.

18. The system of claim 16, further comprising a client that couples to the voice browser through a telephone call.

10

19. The system of claim 18, wherein the terminal object of the voice browser renders the Web page into speech for the client that couples to the voice browser through the telephone call.

15 20. The system of claim 18, wherein the terminal object of the voice browser allows the client to navigate through a Web site based on the speech commands of the client.

21. A data structure to enhance media processing, comprising:
20 a terminal data structure to instantiate terminal objects; and
a speech recognition terminal data structure that extends the terminal data structure.

22. The data structure of claim 21, wherein the speech recognition terminal
25 data structure includes an engine token data structure.

23. The data structure of claim 21, wherein the speech recognition terminal data structure includes an enumeration engine data structure.

24. The data structure of claim 21, wherein the speech recognition terminal data structure includes a speech recognition data structure.

25. The data structure of claim 21, wherein the speech recognition terminal data structure includes a recognition context data structure.

26. A data structure to enhance media processing, comprising:
a terminal data structure to instantiate terminal objects; and
a speech recognition terminal data structure that extends the terminal data structure, wherein the speech recognition terminal data structure includes an engine token data structure.

27. The data structure of claim 26, wherein the engine token data structure includes a method member get engine name for getting a name of a speech recognition engine in a textual form.

28. The data structure of claim 26, wherein the engine token data structure

5 includes a method member get engine token for getting an identifier that identifies
a speech recognition engine.

29. A data structure to enhance media processing, comprising:
a terminal data structure to instantiate terminal objects; and
10 a speech recognition terminal data structure that extends the terminal data
structure, wherein the speech recognition terminal data structure includes an
enumeration engine data structure.

30. The data structure of claim 29, wherein the enumeration engine data
15 structure includes a method member next for getting a next available speech
recognition engine.

31. A data structure to enhance media processing, comprising:
a terminal data structure to instantiate terminal objects; and
20 a speech recognition terminal data structure that extends the terminal data
structure, wherein the speech recognition terminal data structure includes a speech
recognition data structure.

32. The data structure of claim 31, wherein the speech recognition data
25 structure includes a member method enumerate recognition engines for obtaining

5 an indirect reference to a listing of speech recognition engines that are available
for use.

33. The data structure of claim 31, wherein the speech recognition data
structure includes a member method select engine for selecting a speech
10 recognition engine to be used.

34. The data structure of claim 31, wherein the speech recognition data
structure includes a member method get selected engine for retrieving the
currently selected speech recognition engine.

35. The data structure of claim 31, wherein the speech recognition data
structure includes a member method convert extended markup language to
grammar for converting extended markup language text into a compiled grammar
for use with a speech recognition engine.

36. A data structure to enhance media processing, comprising:
a terminal data structure to instantiate terminal objects; and
a speech recognition terminal data structure that extends the terminal data
structure, wherein the speech recognition terminal data structure includes a
25 recognition context data structure.

37. The data structure of claim 36, wherein the recognition context data structure includes a method member initialize for creating a speech recognition context based on a selected speech recognition engine.

10 38. The data structure of claim 36, wherein the recognition context data structure includes a method member shut down for destroying a speech recognition context.

39. The data structure of claim 36, wherein the recognition context data
15 structure includes a method member load grammar for loading a grammar into a recognition context from a source selected from a group consisting of a resource, a memory, and a file.

40. The data structure of claim 36, wherein the recognition context data
20 structure includes a method member unload grammar for unloading a grammar previously loaded into a recognition context.

41. The data structure of claim 36, wherein the recognition context data
structure includes a method member activate grammar for activating a grammar to
25 be used in a speech recognition engine.

42. The data structure of claim 36, wherein the recognition context data structure includes a method member get result for retrieving a speech recognition result.

10 43. The data structure of claim 36, wherein the recognition context data structure includes a method member get hypothesis for retrieving a speech recognition result that is deemed a likely speech recognition result.

44. A method for enhancing media processing, comprising:
15 requesting a speech recognition terminal object;
getting a desired speech recognition engine; and
setting a speech recognition context.

45. The method of claim 44, further comprising selecting a speech recognition
20 terminal object.

46. The method of claim 44, wherein getting includes enumerating a list of available speech recognition engines.

25 47. The method of claim 46, wherein getting includes identifying a desired

5 speech recognition engine from the list of available speech recognition engines.

48. The method of claim 47, wherein getting includes selecting the desired speech recognition engine.

10 49. The method of claim 44, wherein setting includes initializing the speech recognition context.

50. The method of claim 44, wherein setting includes loading a grammar for the speech recognition context.

15

51. The method of claim 44, wherein setting includes activating a grammar for the speech recognition context.

52. The method of claim 44, wherein setting includes setting the speech
20 recognition context to notify a user when a desired event occurs.

~~53.~~ A computer readable medium having instructions stored thereon for causing a computer to perform a method for enhancing media processing, the method comprising:

25 requesting a speech recognition terminal object;

5 getting a desired speech recognition engine; and
setting a speech recognition context.

54. A data structure to enhance media processing, comprising:
a terminal data structure to instantiate terminal objects; and
10 a speech generation terminal data structure that extends the terminal data
structure.

55. The data structure of claim 54, wherein the speech generation terminal data
structure includes voice method members that are selected from a group consisting
15 of a method member set voice for setting a voice to be used for speech generation
and a method member get voice for getting the voice used in speech generation.

56. The data structure of claim 54, wherein the speech generation terminal data
structure includes priority method members that are selected from a group
20 consisting of a method member set priority for setting a priority for a voice and a
method member get priority for getting a priority for a voice, wherein a voice with
a higher priority may interrupt a voice with a lower priority.

57. The data structure of claim 54, wherein the speech generation terminal data
25 structure includes volume method members that are selected from a group

5 consisting of a method member set volume for setting a volume of speech
synthesized by a speech generation engine and a method member get volume for
getting a volume of speech synthesized by a speech generation engine.

58. The data structure of claim 54, wherein the speech generation terminal data
10 structure includes rate method members that are selected from a group consisting
of a method member set rate for setting a rate of speech synthesized by a speech
generation engine and a method member get rate for getting a rate of speech
synthesized by a speech generation engine.

59. The data structure of claim 54, wherein the speech generation terminal data
15 structure includes time out method members that are selected from a group
consisting of a method member set time for setting a time for a speech synthesis to
time out and a method member get time for getting a time for a speech synthesis to
time out.

20 60. The data structure of claim 54, wherein the speech generation terminal data
structure includes a method member speak for synthesizing text to audio.

61. The data structure of claim 54, wherein the speech generation terminal data
25 structure includes a method member get status for getting a status on synthesizing

5 of output audio.

62. The data structure of claim 54, wherein the speech generation terminal data structure includes a method member skip for skipping to a specific point in a text stream.

10

63. The data structure of claim 60, wherein the speech generation terminal data structure includes a method member wait for blocking other executions until the method member speak has been executed to completion.

15 64. The data structure of claim 60, wherein the speech generation terminal data structure includes a method member enumerate voices for obtaining a list of voices for the speech generation engine.

20 65. A data structure to enhance media processing, comprising:
a terminal data structure to instantiate terminal objects; and
a speech generation terminal data structure that extends the terminal data structure, wherein the speech generation terminal data structure includes a method member speak for synthesizing text to audio.

25 66. The data structure of claim 65, wherein the method member speak is

5 receptive to a text stream with voice markup to be synthesized.

67. The data structure of claim 65, wherein the method member speak is receptive to an offset that represents an offset into a text stream where the voice should start speaking.

10

68. The data structure of claim 65, wherein the method member speak is receptive to a speakover flag so as to blend the voice output over any currently playing audio output.

15 69. The data structure of claim 65, wherein the method member speak is receptive to a punctuation flag so as to allow a speech generation engine to speak each punctuation of a text stream.

20 70. A method for enhancing media processing, comprising:
requesting a speech generation terminal object; and
generating a speech.

71. The method of claim 70, wherein generating includes generating the speech from a text stream that includes voice markup.

25

5 72. The method of claim 70, further comprising selecting a voice.

73. The method of claim 72, wherein selecting includes enumerating a list of available voices.

10 74. The method of claim 73, wherein selecting includes identifying a desired voice from the list of available voices.

75. A computer readable medium having instructions stored thereon for causing a computer to perform a method for enhancing media processing, the
15 method comprising:
requesting a speech generation terminal object; and
generating a speech.